

Ejercicio Final Módulo 3

Ejercicio

– Fase 1: Exploración y Limpieza

1. Exploración Inicial:

- ☐ Realiza una exploración inicial de los datos para identificar posibles problemas, como valores nulos, atípicos o datos faltantes en las columnas relevantes.
- ☐ Utiliza funciones de Pandas para obtener información sobre la estructura de los datos, la presencia de valores nulos y estadísticas básicas de las columnas involucradas.
- ☐ Une los dos conjuntos de datos de la forma más eficiente.

2. Limpieza de Datos:

- ☐ Elimina o trata los valores nulos, si los hay, en las columnas clave para asegurar que los datos estén completos.
- ☐ Verifica la consistencia y corrección de los datos para asegurarte de que los datos se presenten de forma coherente.
- ☐ Realiza cualquier ajuste o conversión necesaria en las columnas (por ejemplo, cambiar tipos de datos) para garantizar la adecuación de los datos para el análisis estadístico.

– Fase 2: Visualización

Usando las herramientas de visualización que has aprendido durante este módulo, contesta a las siguientes gráficas usando la mejor gráfica que consideres:

3. **¿Cómo se distribuye la cantidad de vuelos reservados por mes durante el año?**
4. **¿Existe una relación entre la distancia de los vuelos y los puntos acumulados por los clientes?**
5. **¿Cuál es la distribución de los clientes por provincia o estado?**
6. **¿Cómo se compara el salario promedio entre los diferentes niveles educativos de los clientes?**
7. **¿Cuál es la proporción de clientes con diferentes tipos de tarjetas de fidelidad?**
8. **¿Cómo se distribuyen los clientes según su estado civil y género?**

– Fase 3: Evaluación de Diferencias en Reservas de Vuelos por Nivel Educativo

Objetivo del Ejercicio:

Utilizando un conjunto de datos que hemos compartido, se busca evaluar si existen diferencias significativas en el número de vuelos reservados según el nivel educativo de los clientes. Para ello, los pasos que deberas seguir son:

9. Preparación de Datos:

- ☐ Filtra el conjunto de datos para incluir únicamente las columnas relevantes: 'Flights Booked' y 'Education'.

10. Análisis Descriptivo:

- ☐ Agrupa los datos por nivel educativo y calcula estadísticas descriptivas básicas (como el promedio, la desviación estandar, los percentiles) del número de vuelos reservados para cada grupo.

11. Prueba Estadística:

- ☐ Realiza una prueba de A/B testing para determinar si existe una diferencia significativa en el número de vuelos reservados entre los diferentes niveles educativos.

Los Datos

* *Customer Loyalty History.csv

* *Customer Flight Activity.csv